



(11) Publication number:

0 079 739
A2

(12)

EUROPEAN PATENT APPLICATION

(21) Application number: 82305926.6

(22) Date of filing: 08.11.82

 (51) Int. Cl.³: **C 12 N 15/00**
C 12 N 1/00, C 12 P 21/02
C 07 H 21/04, C 07 C 103/52
//C12R1/19, C12R1/865

(30) Priority: 12.11.81 US 320632

 (43) Date of publication of application:
 25.05.83 Bulletin 83/21

 (64) Designated Contracting States:
 BE CH DE FR GB IT LI NL SE

 (71) Applicant: **THE UPJOHN COMPANY**
 301 Henrietta Street
 Kalamazoo, Michigan 49001(US)

 (72) Inventor: **Dugaiczuk, Achilles**
 c/o The Upjohn Company 301 Henrietta Street
 Kalamazoo Michigan 49001(US)

 (74) Representative: **Perry, Robert Edward et al,**
GILL JENNINGS & EVERY 53-64 Chancery Lane
 London WC2A 1HN(GB)

 (54) **Albumin-based nucleotides, their replication and use, and plasmids for use therein.**

 (57) The DNA sequence coding for human serum albumin has been isolated and inserted as two fragments into two novel plasmids which can be replicated in *E. coli*. These novel fragments can be joined to provide a unitary DNA sequence which then can be cloned into a suitable host, e.g. *E. coli*, for the expression of human serum albumin (which is used extensively in medical practice in treating shock conditions).

EP 0 079 739 A2

ALBUMIN-BASED NUCLEOTIDES, THEIR REPLICATION
AND USE, AND PLASMIDS FOR USE THEREIN

This invention relates to nucleotides related to human serum albumin (HSA), their replication and use, and plasmids (and host substances) for use therein.

The gene for serum albumin is regulated in development. On the other hand, serum albumin is synthesised in mammals by the adult liver, and its plateau in adulthood. The embryonic liver and yolk sac, on the other hand, produce predominantly α -fetoprotein, but the synthesis decreases drastically after birth. Recently, 10 Law et al determined the complete sequence of mouse α -fetoprotein mRNA, Nature 291 (1981) 201-205. The structure revealed extensive homology to mammalian serum albumin, indicating that the two proteins are encoded in the same gene family. Similar conclusions have been 15 reached from studies on the α -fetoprotein genes of the rat and the mouse; see Jagodzinski et al, Proc. Natl. Acad. Sci. USA, 78 (1981) 3521-3525, and Gorin et al, J. Biol. Chem. 256 (1981) 1954-1959.

The complete nucleotide sequence of human serum 20 mRNA has been determined from recombinant cDNA clones and from a primer-extended cDNA synthesis on the mRNA template. The sequence comprises 2,078 nucleotides, starting upstream of a potential ribosome binding site in the 5'-untranslated region. It contains all the 25 translated codons and extends into the poly(A) at the 3'-terminus. Part of the translated sequence codes for a hydrophobic prepeptide met-lys-trp-val-thr-phe-ile-ser-leu-leu-phe-leu-phe-ser-ser-ala-tyr-ser, followed by a basic propeptide arg-gly-val-phe-arg-arg. These signal 30 peptides are absent from mature serum albumin and, so far, have not been identified in their nascent state in humans. A remaining 1,755 nucleotides of the translated mRNA sequence code for 585 amino acids which are in agreement, with few exceptions, with the published amino 35 acid data for human serum albumin. The mRNA sequence verifies and refines the repeating homology in the triple-domain structure of the serum albumin molecule.

DETAILED DESCRIPTION OF THE INVENTION

Human serum albumin cDNA is cloned into the PstI site of plasmid pBR322 by the oligo(dG)-oligo(dC) tailing technique. Plasmid DNA was isolated from 97 positive colonies which hybridized to the enriched albumin cDNA probe, and the recombinant plasmid pHA36 was found to contain the largest insert of an albumin cDNA sequence. Its restriction endonuclease map is shown in the drawing, together with a restriction map of the primer-extended plasmid clone pHA206. The latter was obtained in a second transformation experiment after initiating the cDNA synthesis from an internal primer. This primer was a 91 base pairs long DNA fragment, MspI(152)-TaqI(182/3), isolated from pHA36. The two plasmids, pHA36 and pHA206, share 0.15 kb of homologous DNA. Together, they encode the entire sequence for human serum albumin, starting with the CTT codon for leu -10 of the prepeptide and extending into the 3'-untranslated region of poly(A).

Sequence of the Albumin cDNA. The sequence was determined for the most part on both DNA strands to ensure accuracy. All of the restriction sites used to end-label DNA fragments were sequenced across by labeling a neighboring restriction site. The entire nucleotide sequence of the serum albumin mRNA, as determined from the cloned DNA in pHA36, pHA206, and from the primer-extended cDNA at the 5'-terminus of the message, is shown in the following Table 1. The inferred amino acid sequence is also indicated. The mRNA length is 2,078 nucleotides, of which 38 represent the 5'-untranslated region, 54 identify a prepeptide of 18 amino acids, 18 identify a propeptide of 6 amino acids, 1,755 code for the known 585 amino acids of serum albumin, 189 make up the 3'-untranslated region and 24 are the poly(A) sequence. Nucleotides 5 to 15 (-34 to -24) in the 5'-untranslated region (Table 1) are complementary to a 3'-terminal region of eukaryotic 18S RNA [Azad, A.A. and Deacon, N.J. (1980) Nucl. Acids Res. 8, 4365-4376] and thus could represent a ribosome binding site:

(5')...T T^CT C T T C T G T.....albumin mRNA
 (3')...G A G G A A G G C G U C C m₂⁶A m₂⁶A.....18S RNA

The translated portion of the mRNA sequence codes for the signal peptide and the main body of the albumin polypeptide chain. The

signal peptide is composed of a hydrophobic prepeptide of 18 amino acids and a basic propeptide of 6 amino acids (Table 1). Since pre-peptides are removed from nascent secretory proteins (like albumin) in the endoplasmic reticulum, they are seen only in vitro in heterologous
5 translation systems. As yet, they have not been found within cells [Judah, J.D. and Quinn, P.S. (1977) FEBS 11th Mtg., Copenhagen 50, 21-29; and Strauss, A.W., Donohue, A.M., Bennett, C.D., Rodkey, J.A. and Alberts, A.W. (1977) Proc. Natl. Acad. Sci. USA 74, 1358-1362]. This is the first report of the presence and the sequence of a pre-
10 peptide for human serum albumin. As it is with other secretory proteins, the conversion of proalbumin to albumin takes place in the Golgi vesicles, and the enzyme responsible for this cleavage is probably cathepsin B [Judah, J.D. and Quinn, P.S. (1978) Nature 271, 384-385]. This is also a first report on the sequence of the pro-
15 peptide for normal human serum albumin.

At the 3'-end of the message, the putative polyadenylation signal sequence, AATAAA, is located 164 nucleotides downstream from the amino acid termination codon TAA and 16 nucleotides upstream from the beginning of the poly(A) sequence. Another characteristic sequence
20 located near the polyadenylation site has been identified by Benoist, et al. [Benoist, C., O'Hare, K., Breathnach, R. and Chambon, P. (1980) Nucl. Acids Res. 8, 127-142]; the consensus sequence from several mRNAs was concluded as TTTTCACTGC. A similar sequence, TTTTCTCTGT, is located 19 nucleotides upstream from the AATAAA hexanucleotide in the
25 human albumin mRNA (Table 1).

TABLE 1

	-18	p r o	-10		
		Met lys trp val tlu phe lle ser leu leu phe leu phe ser			
		GCTTTTCTTCTGTCAACCCGACAGCCTTTGGCACATG MAG TGC GTA ACC TTT ATT TCC CTT CTT TTT CTC TTT AGC	(30)		
	-1	p r o	-6		
		ser ala tyr ser arg gly val phe arg asp ala his lys ser glu val ala his arg phe lys asp leu qly qlu asn phe lys			
		TGC GCT TAT TCC ACG CGT GTG TTT CGT CCA CAT GCA CAC AAC AGT CAG CTT CCT CAT CGG TTT AAA GAT TTG CCA GAA CAA AAT TTC AAA	(170)		
	21		30	34	40
		ala leu val leu lle ala phe ala gln tyr leu gln gln cys pro phe qlu asp his val lys leu val asn qlu val thr qlu phe ala			
		GCCC TTG GTG TTG ATT GCC TTT GCT CAG TAT CTT CAG CAG TGT CCA TTT CAA CAT CAT GTA AAA TTA GTG AAT GAA GTA ACT CAA TTT GCA	(260)		
	51	53	60	62	70
		lys thr cys val ala asp glu ser ala glu asn cys asp lys ser leu his thr leu phe gly asp lys leu cys thr val ala thr leu			
		AAA ACA TGT GTT CTT CATT CAG TCA GCT GAA AAT TGT GAC AAA TCA CTT CAT ACC CTT TTT CGA CAC AAA TTA TGC ACA GTT CCA ACT CTT	(350)		
	81		90	91	100
		arg glu thr tyr gly glu met ala asp cys cys ala lys gln glu pro qly arg asn qlu cys phe leu qin his lys asp asp asn pro			
		CCT GAA ACC TAT CGT GAA ATG CQT CAC TGC TGT GCA AAA CAA CAA CCT CGG ACA AAT GAA TGC TTC TTG CAA CAC AAA CAT CAC AAC CCA	(440)		
	111		120	124	130
		asn leu pro arg leu val arg pro glu val asp val met cys thr ala phe his asp asn qlu qlu thr phe leu lys lys tyr leu try			
		AAC CTC CCC CCA TTG CTG ACA CCA GAG GTT CAT GTG ATG TGC ACT GCT TTT CAT CAC AAT GAA CAG ACA TTT TTG AAA AAA TAC TTA TAT	(330)		
	141		150	160	168
		glu lle ala arg arg his pro tyr phe tyr ala pro glu leu leu phe phe ala lys arg tyr lys ala ala phe thr qlu cys cys qin			
		GAA ATT GCC ACA ACA CAT CCT TAC TTT TAT GCC CCG GAA CTC CTT TTC TTT GCT AAA AGG TAT AAA GCT CCT TTT ACA CAA TGT TGC CAA	(620)		
	171		180	190	200
		ala ala asp lys ala ala cys leu leu pro lys leu asp glu leu arg asp qlu qly lys ala ser ser ala lys qin arg leu lys cys			
		GCT GCT CAT AAA GCT GCC TGC CTG TTG CCA MAG CTC CAT GAA CTT CGG GAT GAA CGG MAG CCT TCG TCT GCC AAA CAG ACA CTC MAG TGT	(710)		
	201		210	220	230
		ala ser leu gln lys phe gly qlu arg ala phe lys ala trp ala val ala arg leu ser gln arg phe pro lys ala qlu phe ala qlu			
		GCC AGT CTC CAA AAA TTT GGA GAA ACA GCT TTC AAA GCA TGC GCA GTA GCT CGC CTG ACC CAG ACA TTT CCC AAA CCT GAG TTT GCA GAA	(300)		

231 val ser lys leu val thr asp leu thr lys val his thr glu cys oys his gly asp leu leu glu cys ala asp asp arg ala asp leu 260
 GTT TCC AAC TTA GTG ACA CAT CTT ACC AAA GTC CAC ACC GAA TCC TCC CAT GCA CAT CTG CTT GAA TGT CTT CAT CAC AGC GCG CAC CTT (890)

 261 ala lys tyr ile cys glu asn gln asp ser ile ser ser lys leu lys glu cys glu lys pro leu leu glu lys ser his cys ile 289 290
 CCC AAG TAT ATC TGT GAA AAT CAA CAT TCG ATC TCC AGT AAA CTG AAG GAA TCC TGT GAA AAA CCA CTT TGC GAA AAA TCC CAC TCC ATT (980)

 291 ala glu val glu asn asp glu met pro ala asp leu pro ser leu ala ala asp phe val glu ser lys asp val cys lys asn tyr ala 320
 CCC GAA GTG CAA AAT CAT GAG ATG CCT GCT CAC TTG CCT TCA TTA CCT GCT CAT TTT GTT GAA AGT AAG CAT GTT TGC AAA AAC TAT CTT (1070)

 321 glu ala lys asp val phe leu gly met phe leu tyr glu tyr ala arg arg his pro asp tyr ser val val leu leu arg leu ala 350
 CAG GCA AAG CAT GTC TTC TTG GGC ATG TTT TTG TAT GAA TAT GCA ACA AGC CAT CCT CAT TAC TCT GTC GTG CTG CTG ACA CTT GCG (1160)

 351 lys thr tyr glu thr thr leu glu lys cys cys ala ala asp pro his glu cys tyr ala lys val phe asp glu phe lys pro leu 380
 AAG ACA TAT GAA ACC ACT CTA GAG AAG TCC TGT GCT GCA CAT CCT CAT GAA TGC TAT GCC AAA GTG TTC GAT GAA TTT AAA CCT CCT (1250)

 381 val glu glu pro gln asn leu ile lys gln asn cys glu leu phe glu leu qly qly tyr lys phe qln asn ala leu leu val arg 410
 CTG CAA GAG CCT CAG AAT TTA ATC AAA CAA AAT TGT GAG CTT TTT CAG CAG CTT GCA GAG TAC AAA TTC CAG AAT GCG CTG TTA GTT CTT (1340)

 411 tyr thr lys lys val pro gln val ser thr pro thr leu val glu val ser arg asn leu qly lys val qly ser lys cys cys lys his 440
 TAC ACC AAG AAA GTA CCC CAA GTG TCA ACT CCA ACT CTT GTA CAG GTC TCA AGA AAC CTA GCA AAA GTG GCG AGC AAA TGT TGT AAA CAT (1430)

 441 pro glu ala lys arg met pro cys ala glu asp tyr leu ser val val leu asn gln leu cys val leu his glu lys thr pro val ser 470
 CCT GAA GCA AAA ACA ATG CCC TGT GCA GAA GAC TAT CTA TCC CTC CTC CAG TTA TGT TGT GTC TTG CAT GAG AAA ACC CCA GTA AGT (1520)

 471 asp arg val thr lys cys cys thr glu ser leu val asn arg arg pro cys phe ser ala leu glu val asp glu thr tyr val pro lys 500
 CAC ACA CTC ACC AAA TGC TGC ACA GAA TCC TTG CTC AAC ACC CCA CCA TCC TTT TCA CCT CTG CAG GAT GAA GTC GAT GAA ACA TAC GTT CCC AAA (1610)

 501 glu phe asn ala glu thr phe thr phe his ala asp ile cys thr leu ser glu lys glu arg qln ile lys lys qln thr ala leu val 530
 CAG TTT AAT GCT GAA ACA TTC ACC TTC CAT GCA GAT ATA TCC ACA CTT TCT CAG AAG GAG ACA CAA ATC AAG AAA CAA ACT GCA CTT CTT (1700)

5 558 559 560
glu leu val lys his lys pro lys ala thr lys glu gln leu lys ala val met asp phe ala ala phe val glu lys cys cys lys
GAG CTC GTG AAA CAC AAG CCC AAG CCA ACA AAA GAG CAA CTG AAA GCT GTT ATG CAT CAT TTC CCT GCT TTT GTA GAG AAG TGC TGC AAG (1790)

10 550
561 567 570
ala asp asp lys glu thr oys phe ala glu qly lys lys leu val ala ala ser gln ala ala leu qly leu ter
CCT GAC GAT AAG CAG ACC TGC TTT GCC GAG CAG GGT AAA AAA CTT GTT GCT CCA AGT CAA GCT GCC TTA GGC TTA TAA CATCACATTAAAG (1883)

15 580
ter ter
CATCTCAGCCTACCATGAGATAAGAGAAAGAAATCAGATCAAAAGCTTATTCATCTGTTTTCTTTTCGTTGCTAAGCCACACCCCTGCTAAACATATAATTCTTTAA (2002)

20
TCATTTGCCCTCTTTCTCTGCTCTCAATTATAATAAAATGGAAGATCTAA..... 20AA (2078)

25 540
30 560
35 531

Following are examples which illustrate procedures, ~~including the best mode~~, for practicing the invention. These examples should not be construed as limiting. All percentages are by weight and all solvent mixture proportions are by volume unless otherwise noted.

5 Example 1 Isolation of Messenger RNA

Human liver mRNA was obtained following the procedure of Chirgwin, et al [Chirgwin, J.M., Przybyla, A.E., MacDonald, R.J. and Rutter, W.J. (1979) Biochemistry 18, 5294-5299]. Immunoprecipitation of albumin containing polysomes was performed according to Taylor and
10 Tse [Taylor, J.M. and Tse, T.P.H. (1976) J. Biol. Chem. 251, 7461-7467]. In vitro translation of mRNA was carried out in a reticulocyte cell-free system, following the instruction of the manufacturer (New England Nuclear). The translation products were separated electrophoretically according to Laemmli [Laemmli, J.K. (1970) Nature 227, 680-685].
15

Example 2 Cloning Procedures

Double stranded cDNA was synthesized as described previously [Law, S., Tamaoki, T., Kreuzaler, F. and Dugaiczky, A. (1980) Gene 10, 53-61]. It was annealed to PstI-linearized pBR322 DNA [Rolivar, F.,
20 Rodriguez, R.L., Greene, P.J., Betlach, M.C., Heyneker, H.L., Boyer, H.W., Crossa, J.H. and Falkow, S. (1977) Gene 2, 95-113] that had been tailed with 15 dG residues/3'-terminus [Dugaiczky, A., Robberson, D.L. and Ullrich, A. (1980) Biochemistry 19, 5869-5873]. The annealed DNA was used to transform E. coli strain RR1, as detailed previously [Law, S., et al., Ibid.]. The albumin clones were selected using the colony
25 hybridization method of Grunstein and Hogness [Grunstein, M. and Hogness, D.S. (1975) Proc. Natl. Acad. Sci. USA 72, 3961-3965], with [³²P]-labeled cDNA synthesized with the immunoprecipitated polysomal mRNA as template.

30 As shown in Example 5, plasmids pHA36 and pHA206 were deposited in E. coli HB101 hosts. The plasmids were obtained from E. coli RR1 hosts, described in this example, and transformed into E. coli HR101 by standard procedures well known to those of ordinary skill in this art. The E. coli RR1 hosts were lysed and then centrifuged to
35 separate the chromosomal DNA, cell DNA and plasmid DNA. The plasmid DNA, remaining in the supernatant, is precipitated with ethanol and the precipitate is resuspended in buffer, e.g., TCM (10mM Tris·HCl, pH 8.0, 10 mM CaCl₂, 10 mM MgCl₂). The cells for transformation are

prepared as follows: 120 ml of L-broth (1% tryptone, 0.5% yeast extract, 0.5% NaCl) are inoculated with an 18 hour culture of HB101 NRRL B-11371 and grown to an optical density of 0.6 at 600 nm. Cells are washed in cold 100 mM NaCl and resuspended for 15 minutes in 20 ml chilled 50 mM CaCl₂. Bacteria are then concentrated to one-tenth of this volume in CaCl₂ and mixed 2:1 (v:v) with annealed plasmid DNA, prepared as described above. After chilling the cell-DNA mixture for 15 minutes, it is heat shocked at 42°C for 2 minutes, then allowed to equilibrate at room temperature for ten minutes before addition of L-broth 10 times the volume of the cell-DNA suspension. Transformed cells are incubated in broth at 37°C for one hour before inoculating selective media (L-agar plus 10 µg/ml tetracycline) with 200 µl/plate. Plates are incubated at 37°C for 48 hours to allow the growth of transformants.

15 Example 3 Mapping of Restriction Endonuclease Sites

Restriction endonucleases were obtained from Bethesda Research Laboratories and New England Biolabs and were used according to the manufacturers' instructions. The digested DNA fragments were analyzed electrophoretically on agarose [Helling, R.R., Goodman, H.M. and Boyer, H.W. (1974) J. Virol. 14, 1235-1244] or acrylamide [Dingman, C., Fisher, M.P. and Kakefuda, T. (1972) Biochemistry 11, 1242-1250] gels.

20 Example 4 DNA Sequencing

DNA fragments were dephosphorylated with bacterial alkaline phosphatase (Worthington) and labeled at the 5'-ends with polynucleotide kinase (Boehringer-Mannheim) and γ [³²P]ATP. Following digestion with a second restriction endonuclease and electrophoretic separation of the fragments, DNA sequence determination was done according to the procedure of Maxam and Gilbert [Maxam, A. and Gilbert, W. (1980) Methods Enzym. 65, 499-560] and the degradation products were separated electrophoretically on 0.4 mm acrylamide gels as described by Sanger and Coulson [Sanger, F. and Coulson, R. (1978) FEBS Letters 87, 107-110].

30 Example 5 Recombinant Plasmids pHA36 and pHA206

As disclosed in Example 2, albumin clones were selected by hybridizing to the enriched albumin cDNA probe. Plasmid pHA36 contained the largest insert of an albumin cDNA sequence. Both plasmids pHA36 and pHA206 have been deposited in a viable E. coli host in the

permanent collection of the Northern Regional Research Laboratory (NRRL), U.S. Department of Agriculture, Peoria, Illinois, U.S.A. Their accession numbers in this repository are as follows:

HB101(pHA36) - NRRL B-12551

5

HB101(pHA206) - NRRL B-12550

E. coli HB101 is a known and widely available host microbe. Its NRRL accession number is NRRL B-11371.

NRRL B-12550 and NRRL B-12551 are available to the public, ~~upon the grant of a patent. It should be understood that the availability~~
10 of these deposits does not constitute a license to practice the subject invention in derogation of patent rights granted with the subject instrument by governmental action.

E. coli RR1 and E. coli HB101 are known and widely available host microbes. Their NRRL accession numbers are NRRL B-12186 and NRRL
15 B-11371, respectively.

pBR322 is a well known and widely available plasmid. It can be obtained from the following host deposit by standard procedures:

NRRL B-12014 - E. coli RR1 (pBR322).

YEp6 is a well known and widely available yeast episomal plasmid.
20 It can be obtained from the following host deposit by standard procedures:

E. coli HB101 (YEp6) - NRRL B-12093.

Example 6 Assembly of the Serum Albumin Gene

Assembling the pieces together is a straightforward task of re-
25 striction enzymology. There is only one MspI site in the overlapping DNA sequence of the two cDNA clones. Two enzymatic steps of (i) MspI digestion of the two DNAs, followed by (ii) the use of ligase, an enzyme that seals DNA fragments, will give the desired product. Although two other undesired DNA species will also be obtained in the
30 course of this recombination reaction, both of them will differ substantially in size. Thus, separation and isolation of the desired DNA species will be achieved.

The assembled DNA clone can be used to transform two types of cells:

- 35 (a) Escherichia coli
 (b) Saccharomyces cerevisiae

(a) The vector of choice is plasmid pBR322, the same that has

been successfully used for cloning of the two fragmented pieces of the serum albumin cDNA.

(b) In order to transform yeast with the serum albumin structural gene sequence, the DNA must be inserted into one of the existing yeast plasmid vectors. This can be accomplished by taking advantage of the fact that several restriction endonuclease recognition sequences are absent from the cloned serum albumin DNA. Synthetic EcoRI DNA linkers can be ligated to the DNA fragment containing the serum albumin sequence followed by insertion (ligation) into one of the yeast plasmid vectors, e.g., YEp6, at the Eco RI cloning site. The fused chimeric plasmid can be used to transform yeast according to an established procedure [Hinnen, A., Hicks, J.R. and Fink, G.R. (1978) Proc. Natl. Acad. Sci. USA, 75, 1929]. YEp6 can be obtained from the NRRL repository, as disclosed supra.

15 Example 7 Expression of the Serum Albumin Gene

The main body of the structural gene will be transcribed by the E. coli or yeast enzymes. If little or no albumin is produced with the selected host, then an Escherichia coli promoter DNA sequence carrying an initiation codon, i.e., ATG, can be ligated at the beginning of the serum albumin structural gene. Such elements are known and available, e.g., lac promoter used for the expression of human interferon gene in E. coli [Proc. Natl. Acad. Sci. 77, 5230 (1980)]; source of promoter DNA [Proc. Natl. Acad. Sci. 76, 760 (1979)]. Also, see Nature, Vol. 281, October 18, 1979. It has already been documented that such Escherichia coli promoter sequences function well in the expression of foreign genes in Escherichia coli [Mercereau-Pujalon, O., Royal, A., Cami, B., Garapin, A., Krust, A., Gannon, I. and Kourilsky, P. (1978) Nature 275, 505; and Goeddel, D.V., Kleid, D.G., Bolivar, F., Heyneker, H.L., Yansura, D.G., Grea, R., Hirose, T., Kraszewski, A., Itakura, K., and Riggs, A. (1979) Natl. Acad. Sci. USA 76, 106]. For expression in yeast, see Rose, M., Casadaban, M.J. and Botstein, D. (1981) Proc. Natl. Acad. Sci. USA 78, 2460 and 4466.

30 Example 8 Screening of Clones Producing Albumin

Immunological methods can be used to detect small amounts of albumin made in a bacterium. Flat disks of flexible polyvinyl are coated with the IgG fraction from an immune serum and the disks are pressed onto an agar plate so that antigen released from an in situ lysed microbial colony can bind to the fixed antibody. The plastic

disk is then incubated with the same total IgG fraction labeled with radioactive iodine so that other determinants on the bound antigen can in turn bind the iodinated antibody. Radioactive areas on the disk expose X-ray film during autoradiography and thus identify colonies
5 producing the protein which is being screened for. Detailed protocols of this procedure have been published [Broome, S. and Gilbert, W. (1978) Proc. Natl. Acad. Sci. USA, 75, 2746]. The purification of human serum albumin can be accomplished by using procedures well known in the art. For example, procedures disclosed in a chapter by T.
10 Peters: Purification and Properties of Serum Albumin, in: The Plasma Proteins, Putnam, Ed. Academic Press, New York, 1975, can be used.

The work described herein was all done in conformity with physical and biological containment requirements specified in the NIH Guidelines.

15

20

25

30

35

CLAIMS

1. Plasmid pHA36, having a restriction endonuclease pattern as shown in the drawing.

5

2. Plasmid pHA206, having a restriction endonuclease pattern as shown in the drawing.

3. E. coli HB101 (pHA36) having the deposit accession number
10 NRRL B-12551.

4. E. coli HB101 (pHA206) having the deposit accession number
NRRL B-12550.

15 5. A microorganism modified to contain a nucleotide sequence coding for the amino acid sequence of human serum albumin; said nucleotide sequence is as follows:

20

25

30

35

35
 30
 25
 20
 15
 10
 5

-1 -6 p r o -1 1
 ser ala tyr ser arg gly val phe arg arg asp ala his lys ser glu val ala his arg phe lys asp leu gly glu asn phe lys
 TCG CCT TAT TCC ACC CGT GTG TTT CGT CCA CAT GCA CAC MAG AGT CAG GTT GCT CAT CGG TTT AAA GAT TTG GCA GAA AAT TTC AAA (170)

21
 30 34 40 50
 ala leu val leu ile ala phe ala gln tyr leu gln gln cys pro phe glu asp his val lys leu val asn glu val thr glu phe ala
 GCC TTG GTG TTG ATT CCC TTT CCT CAG TAT CTT CAG CAG TGT CCA TTT GAA CAT CAT GTA AAA TTA GTG AAT GAA GTA ACT CAA TTT GCA (260)

51 53 60 62 70 75 80
 lys thr oys val ala asp glu ser ala glu asn cys asp lys ser leu his thr leu phe gly asp lys leu cys thr val ala thr leu
 AAA ACA TGT GTT CCT CAT CAG TCA CCT GAA AAT TGT CAC AAA TCA CTT CAT ACC CTT TTT CGA CAC AAA TTA TCC ACA GTT GCA ACT CTT (350)

81 90 91 100 101 110
 arg glu thr tyr gly glu met ala asp cys cys ala lys gln glu pro gly arg asn glu cys phe leu gln his lys asp asp asn pro
 CGT GAA ACC TAT CGT GAA ATG CCT CAC TCC TGT GCA AAA CAA GAA CCT GGG AGA AAT GAA TGC TTC CAA CAC AAA GAT CAC AAC CCA (440)

111 120 124 130 140
 asn leu pro arg leu val arg pro glu val asp val met cys thr ala phe his asp asn glu thr phe leu lys lys tyr leu try
 AAC CTC CCC CCA TTG GTG ACA CCA CAG GTT GAT GTG ATG TCC ACT GCT TTT CAT GAC AAT GAA CAG ACA TTT TTG AAA AAA TAC TTA TAT (330)

141 150 160 168 169 170
 glu ile ala arg arg his pro tyr phe tyr ala pro glu leu leu phe phe ala lys arg tyr lys ala ala phe thr glu cys cys gln
 CAA ATT GCC ACA ACA CAT CCT TAC TTT TAT GCC CCC GAA CTC CTT TTT CCT AA AGC TAT AAA CCT CCT TTT ACA GAA TGT TCC CAA (620)

171 177 180 190 200
 ala ala asp lys ala ala cys leu leu pro lys leu asp glu leu arg asp glu gly lys ala ser ser ala lys gln arg leu lys cys
 GCT GCT CAT AAA GCT GCC TCC CTG TTG CCA MAG CTC CAT GAA CTT CGG GAT GAA GGG MAG GCT TCG TCT GCC AAA CAG ACA CTC AAG TGT (710)

201 210 220 230
 ala ser leu gln lys phe gly glu arg ala phe lys ala trp ala val ala arg leu ser gln arg phe pro lys ala glu phe ala glu
 GCC AGT CTC CAA AAA TTT GCA GAA ACA GCT TTC AAA GCA TGG GCA GTA GCT CCC CTG AGC CAG ACA TTT CCC AAA GCT GAG TTT GCA GAA (300)

231 val ser lys leu val thr asp leu thr lys val his thr glu cys his gly asp leu leu glu cys ala asp asp arg ala asp leu
 GTT TCC AAG TTA GTG ACA GAT CTT ACC AAA GTC CAC ACG GAA TGC TCC CAT CGA GAT CTG CTT CAA TGT CCT CAT CAC AGG CCG CAC CTT (890)

261 ala lys tyr ile cys glu asn gln asp ser ile ser ser lys leu lys glu cys glu lys pro leu leu glu lys ser his cys ile
 GCC AAG TAT ATC TGT GAA AAT CAA GAT TCC ATC TCC ACT AAA CTG AAG GAA TCC TGT CAA AAA CCT CTG TTG GAA AAA TCC CAC TGC ATT (980)

291 ala glu val glu asn asp glu met pro ala asp leu pro ser leu ala ala asp phe val glu ser lys asp val oys lys asn tyr ala
 GCC GAA GTG GAA AAT GAT CAG ATG CCT GCT CAC TTG CCT TCA TTA CCT GCT CAT TTT GTT GAA AGT AAG CAT GTT TCC AAA AAC TAT CCT (1070)

321 glu ala lys asp val phe leu gly met phe leu tyr glu tyr ala arg arg his pro asp tyr ser val val leu leu leu arg leu ala
 CAG GCA AAG GAT GTC TTC TTG GCC ATG TTT TTG TAT GAA TAT GCA ACA AGG CAT CCT GAT TAC TCT GTC GTG CTG CTG ACA CTT GCC (1160)

351 lys thr tyr glu thr thr leu glu lys oys ala ala asp pro his glu cys tyr ala lys val phe asp glu phe lys pro leu
 AAG ACA TAT CAA ACC ACT CTA CAG AAG TCC TGT GCT GCA CAT CCT CAT GAA TCC TAT GCC AAA GTG TTC CAT GAA TTT AAA CCT CCT (1250)

381 val glu glu pro gln asn leu ile lys gln asn cys glu leu phe glu qln leu gly glu tyr lys phe gln asn ala leu leu val arg
 GTG CAA CAG CCT CAG AAT TTA ATC AAA CAA AAT TGT CAG CTT TTT CAG CAG CTT GCA CAG TAC AAA TTC CAG AAT GCC CTG TTA GTT CGT (1340)

411 tyr thr lys lys val pro gln val ser thr pro thr leu val glu val ser arg asn leu gly lys val qly ser lys cys oys lys his
 TAC ACC AAG AAA GTA CCC CAA GTG TCA ACT CCA ACT CTT GTA CAG GTC TCA ACA MAC CTA GCA AAA GTG GCC AGC AAA TGT TGT AAA CAT (1430)

441 pro glu ala lys arg met pro oys ala glu asp tyr leu ser val val leu asn gln leu oys val leu his glu lys thr pro val ser
 CCT GAA GCA AAA AGA ATG CCC TGT GCA GAA GAC TAT CTA TCC GTC CTC AAC CAG TTA TGT GTC TTG CAT GAG AAA ACC CCA GTA AGT (1520)

471 asp arg val thr lys oys cys thr glu ser leu val asn arg arg pro cys phe ser ala leu glu val asp glu thr tyr val pro lys
 CAC ACA GTC ACC AAA TGC TGC ACA GAA TCC TTG GTG MAC AGG CGA CCA TCC TTT TCA CCT CTG CAA GTC CAT GAA ACA TAC GTT CCC AAA (1610)

501 glu phe asn ala glu thr phe thr phe his ala asp ile cys thr leu ser glu lys glu arg qln ile lys lys qln thr ala leu val
 CAG TTT AAT GCT GAA ACA TTC ACC TTC CAT GCA GAT ATA TCC ACA CTT TCT GAG GAG AAG GAG ACA CAA AAC AAA CCA ACT GCA CTT GTT (1700)

5
10
15
20
25
30
35

531
glu leu val lys his lys pro lys ala thr lys glu gln leu lys ala val met asp phe ala ala phe val glu lys cys lys
GAG CTC GTG AAA CAC MAG CCC MAG CCA ACA AAA CAG CAA CTG AAA CCT GTT ATG CAT GAT TTC GCT TTT GTA GAG AAG TGC AAG (1790)

540
550
558 559 560

561
ala asp asp lys glu thr cys phe ala glu gln lys lys leu val ala ala ser gln ala ala leu qly leu ter
CCT GAC GAT MAG GAG ACC TGC TTT GCC GAG GAG GGT AAA AAA CTT GTT CCT GCA AGT CAA GCT GCC TTA TAA CATCACATTAAAG (1883)

567
570
580

ter ter

CATCTCAGCCTACCATGACATAAGACAGAAATGACATCAAAAGCTTATTCATCTGTTTTCTTTTCGTTGGTGAAGCCACACCCCTGCTCTATAAAACATAAATTTCTTAA (2002)

TCATTTGCCCTCTTTCTCTGCTTCAATTATAAATAAATGGAAGCATCTAA..... 20AA (2078)

6. Nucleotide sequence of the cDNA of human serum albumin, said nucleotide sequence is as follows:

```

5
10
15
20
25
30
35
1 10 20
asp ala his lys ser glu val ala his arg phe lys asp leu qly glu qiu asn phe lys
GAT GCA CAC AAG AGT CAG GTT CCT CAT CCG TTT AAA GAT TTG CCA GAA GAA AAT TTC AAA (170)

30 40 50
ala leu val leu ile ala phe ala gln tyr leu gln gln oys pro phe glu asp his val lys leu val asn glu val thr qiu phe ala
CCC TTG GTG TTG ATT GCC TTT GCT CAG TAT CTT CAG CAG TGT CCA TTT GAA CAT CAT GTA AAA TTA GTG AAT GAA GTA ACT CAA TTT GCA (260)

51 60 70 75 80
lys thr cys val ala asp glu ser ala glu asn oys asp lys ser leu his thr leu phe qly asp lys leu cys thr val ala thr leu
AAA ACA TGT GTT CCT CAT CAG TCA GCT GAA AAT TGT CAC AAA TCA CTT CAT ACC CTT TTT CCA CAC AAA TTA TGC ACA GTT GCA ACT CTT (350)

81 90 91 100 101
arg glu thr tyr gly glu met ala asp oys oys ala lys gln glu pro gly arg asn glu oys phe leu qln his lys asp asp asn pro
CGT GAA ACC TAT CGT GAA ATG CCT GAC TCC TGT CCA AAA CAA CAA CCT CCG ACA AAT GAA TGC TTC TTG CAA CAC AAA GAT GAC AAC CCA (440)

111 120 124 130 140
asn leu pro arg leu val arg pro glu val asp val met cys thr ala phe his asp asn glu qiu thr phe leu lys lys tyr leu try
AAC CTC CCC CGA TTG GTG ACA CCA CAG GTT GAT GTG ATG TGC ACT CCT TTT CAT CAC AAT GAA CAG ACA TTT TTG AAA AAA TAC TTA TAT (330)

141 150 160 168 169 170
glu ile ala arg arg his pro tyr phe tyr ala pro glu leu leu phe phe ala lys arg tyr lys ala ala phe thr glu oys cys qln
GAA ATT GCC ACA ACA CAT CCT TAC TTT TAT TAT GCC CCG GAA CTC CTT TTT CCT AAA AGG TAT AAA GCT CCT TTT ACA GAA TGT TGC CAA (620)

171 177 180 200
ala ala asp lys ala ala oys leu leu pro lys leu asp glu leu arg asp qiu qly lys ala ser ser ala lys qln arg leu lys oys
CCT CCT GAT AAA GCT GCC TCC CTG TTG CCA AAG CTC GAT GAA CTT CCG CAT GAA CCG AAG CCT TCG TCT GCC AAA CAG ACA CTC AAG TGT (710)

201 210 220 230
ala ser leu gln lys phe gly glu arg ala phe lys ala trp ala val ala arg leu ser gln arg phe pro lys ala qiu phe ala qiu
CCC AGT CTC CAA AAA TTT GCA GAA ACA GCT TTC AAA GCA TGG CCA GTA CCT CCG CTC AGC CAG ACA TTT CCC AAA CCT CAG TTT GCA CAA (300)

```


231	240	245 246	250	253	260
val ser lys leu val thr asp leu thr lys val his thr glu cys cys his gly asp leu leu glu cys ala asp asp arg ala asp leu					
CTT TCC AGC TTA GTG ACA CAT CTT ACC AAA GTC CAC ACG GAA TGC TGC CAT GCA CAT CTG CTT GAA TGT GCT GAT GAC ACG GCG GAC CTT (890)					
261	265	270	278 279 280		289 290
ala lys tyr ile cys glu asn gln asp ser ile ser ser lys leu lys glu cys cys qiu lys pro leu leu glu lys ser his cys ile					
GCC AGC TAT ATC TGT CAA AAT CAA CAT TCG ATC TCC AGT AAA CTG AGC GAA TGC TGT GAA AAA CCT CTG TTC GAA AAA TCC CAC TCC ATT (980)					
291	300	310	316	320	
ala glu val glu asn asp glu met pro ala asp leu pro ser leu ala ala asp phe val glu ser lys asp val cys lys asn tyr ala					
CCC CAA CTG CAA AAT CAT CAG ATG CCT GCT CAC TTG CCT TCA TTA GCT GAT TTT GTT CAA AGT MAG CAT GTT TGC AAA AAC TAT GCT (1070)					
321	330	340	350		
glu ala lys asp val phe leu gly met phe leu tyr glu tyr ala arg arg his pro asp tyr ser val val leu leu leu arg leu ala					
GAG CCA MAG CAT GTC TTC GGC ATG TTT TTG TAT GAA TAT GCA ACA AGC CAT CCT CAT TAC TCT GTC CTG CTG ACA CTT GCT (1160)					
351	360 361	369 370	380		
lys thr tyr glu thr thr leu glu lys cys cys ala ala asp pro his qiu cys tyr ala lys val phe asp glu phe lys pro leu					
AAG ACA TAT CAA ACC ACT CTA CAG AAG TGC TGT GCT GCA CAT CCT CAT CAA TCC TAT GCT AAA GTG TTC CAT GAA TTT AAA CCT CCT (1250)					
381	390 392	400	410		
val glu glu pro gln asn leu ile lys gln asn cys glu leu phe qiu qiu leu gly glu tyr lys phe qiu asn ala leu leu val arg					
GTG CAA CAG CCT CAG AAT TTA ATC AAA CAA AAT TGT CAG CTT TTT CAG CAG CTT GCA CAG TAC AAA TTC CAG AAT GCG CTG TTA GTT GCT (1340)					
411	420	430	437 438	440	
tyr thr lys lys val pro gln val ser thr pro thr leu val glu val ser arg asn leu qly lys val qly ser lys cys cys lys his					
TAC ACC AAG AAA GTA CCC CAA GTG TCA ACT CCA ACT CTT GTA CAG GTC TCA ACA AAC CTA CGA AAA GTG GCG ACC AAA TGT TGT AAA CAT (1430)					
441	448 450	460 461	470		
pro glu ala lys arg met pro cys ala glu asp tyr leu ser val val leu asn gln leu cys val leu his qiu lys thr pro val ser					
CCT CAA CCA AAA ACA ATG CCC TGT GCA GAA GAC TAT CTA TCC GTG GTC CTG AAC CAG TTA TGT GTG TTC CAT GAG AAA ACG CCA GTA AGT (1520)					
471	476 477	490	500		
asp arg val thr lys cys cys thr glu ser leu val asn arg arg pro cys phe ser ala leu glu val asp qiu thr tyr val pro lys					
GAC ACA GTC ACC AAA TGC TGC ACA GAA TCC TTG GTG AAC ACG CCA CCA TGC TTT TCA GCT CTG GAA GTC GAT GAA ACA TAC GTT CCC AAA (1610)					
501	510 514	520	530		
glu phe asn ala glu thr phe thr phe his ala asp ile cys thr leu ser qiu lys glu arg qiu ile lys lys qiu thr ala leu val					
CAG TTT AAT GCT GAA ACA TTC ACC TTC CAT GCA GAT ATA TCC ACA CTT TCT GAG AAG GAG ACA CAA ATC AAG AAA CAA ACT GCA CTT GTT (1700)					

ter
ter

7. Nucleotide sequence coding for the prepeptide of human serum albumin, said nucleotide sequence is as follows:

5

10

15

20

25

30

35

Met lys trp val tlu phe ile ser leu leu phe leu phe ser
GCC TTT TCT CTG TCA ACC CCA CAG CCG CTT TGG CACA ATG AAG TGG GTA ACC TTT ATT TCC CTT CTT TTT CTC TTT ACC (30)

-18 p r o -10

-1 -6 p r o -1

ser ala tyr ser arg gly val phe arg arg
TCC GCT TAT TCC ACC GGT GTC TTT CCG CGA

8. Nucleotide sequence coding for pro human serum albumin, said nucleotide sequence is as follows:

```

5
20
10
15
20
25
30
35
-6 p r o -1 1
arg gly val phe arg arg asp ala his lys ser glu val ala his arg phe lys asp leu gly glu asn phe lys
AAG CGT GTC TTT CGT CCA CAT CCA CAC MAG AGT CAG GTT GCT CAT CGG TTT AAA GAT TTG CCA GAA GAA AAT TTC AAA (170)

21 ala leu val leu ile ala phe ala gln tyr leu gln gln oys pro phe glu asp his val lys leu val asn glu val thr glu phe ala
GCC TTG GTG TTG ATT GCC TTT GCT CAG TAT CTT CAG CAG TGT CCA TTT GAA CAT CAT GTA AAA TTA GTG AAT CAA GTA ACT CAA TTT GCA (260)

51 53 60 62 70 75 80
lys thr oys val ala asp glu ser ala glu asn oys asp lys ser leu his thr leu phe gly asp lys leu oys thr val ala thr leu
AAA ACA TGT GTT CCT CAT CAG TCA CCT CAA AAT TGT GAC AAA TCA CTT CAT ACC CTT TTT CGA GAC AAA TTA TCC ACA GTT CCA ACT CTT (350)

81 90 91 100 101
arg glu thr tyr gly glu met ala asp oys oys ala lys gln glu pro gly arg asn glu oys phe leu qln his lys asp asp asn pro
CGT CAA ACC TAT CGT CAA ATG CCT CAC TGC TGT GCA AAA CAA GAA CCT GCG ACA AAT CAA TGC TTC TGC CAA CAC AAA GAT CAC AAC CCA (440)

111 120 124 130 140
asn leu pro arg leu val arg pro glu val asp val met oys thr ala phe his asp asn glu thr phe leu lys lys tyr leu try
AAC CTC CCC CCA TTG GTG ACA CCA CAG GTT CAT GTG ATG TGC ACT GCT TTT CAT GAC AAT GAA GAG ACA TTT TTG AAA AAA TAC TTA TAT (530)

141 150 160 168 169 170
glu ile ala arg arg his pro tyr phe tyr ala pro glu leu leu phe phe ala lys arg tyr lys ala ala phe thr glu oys oys qln
CAA ATT GCC ACA ACA CAT CCT TAT TAT GCC CCC GAA CTC CTT TTC TTT GCT AAA AGG TAT AAA GCT GCT TTT ACA GAA TGT TGC CAA (620)

171 177 180 190 200
ala ala asp lys ala ala oys leu leu pro lys leu asp glu leu arg asp glu gly lys ala ser ser ala lys qln arg leu lys oys
CCT GCT CAT AAA GCT GCC TGC CTG TTG CCA MAG CTC CAT GAA CTT CCG CAT GAA CCG MAG GCT TCG TCT GCC AAA CAG ACA CTC MAG TGT (710)

201 210 220 230
ala ser leu gln lys phe gly glu arg ala phe lys ala trp ala val ala arg leu ser gln arg phe pro lys ala glu phe ala glu
GCC AGT CTC CAA AAA TTT GCA GAA ACA GCT TTC AAA GCA TGG GCA GTA GCT CCC CTG ACC CAG ACA TTT CCC AAA GCT GAG TTT GCA GAA (800)

```


231 val ser lys leu val thr asp leu thr lys val his thr glu cys cys his gly asp leu leu glu cys ala asp asp arg ala asp leu 260
 GTT TCC AAG TTA GTG ACA GAT CTT ACC AAA GTC CAC ACG GAA TCC TCC CAT GGA GAT CTG CTT GAA TGT GCT GAT CAC ACG GCG GAC CTT (890)

261 ala lys tyr ile cys glu asn gln asp ser ile ser ser lys leu lys glu cys cys glu lys pro leu leu glu lys ser his cys ile 289 290
 CCC AAG TAT ATC TGT GAA AAT CAA GAT TCG ATC TCC AGT AAA CTG AAG GAA TCC TGT GAA AAA CCT CTG TTC GAA AAA TCC CAC TGC ATT (980)

291 ala glu val glu asn asp glu met pro ala asp leu pro ser leu ala ala asp phe val glu ser lys asp val cys lys asn tyr ala 320
 GCC GAA GTG GAA AAT GAT CAG ATG CCT GCT GAT TTA CCT GCT CAT TTT GGT GAA AGT AAG GAT GTT TCC AAA AAC TAT GCT (1070)

321 glu ala lys asp val phe leu gly met phe leu tyr glu tyr ala arg arg his pro asp tyr ser val val leu leu leu arg leu ala 350
 CAG CCA AAG CAT GTC TTC TCG GCC ATG TTT TTG TAT GAA TAT CCA ACA ACG CAT CCT GAT TAC TCT GTC CTG CTG CAC GAT CTT GCC (1160)

351 lys thr tyr glu thr thr leu glu lys cys cys ala ala asp pro his glu cys tyr ala lys val phe asp glu phe lys pro leu 380
 AAG ACA TAT GAA ACC ACT CTA GAG AAG TCC TGT GCT GCT CCA CAT CCT CCA GAG TAC TCC TAT GGC AAA GTG TTC CAT GAA TTT AAA CCT CCT (1250)

381 val glu glu pro gln asn leu ile lys gln asn cys glu leu phe glu gln leu qly glu tyr lys phe gln asn ala leu leu val arg 410
 GTG GAA CAG CCT CAG AAT TTA ATC AAA CAA AAT TGT CAG CTT TTT CAG CAG CTT GCA GAG TAC AAA TTC CAG AAT GCG CTG TTA GTT CGT (1340)

411 tyr thr lys lys val pro gln val ser thr pro thr leu val glu val ser arg asn leu qly lys val qly ser lys cys cys lys his 440
 TAC ACC AAG AAA GTA CCC CAA GTG TCA ACT CCA ACT CTT GTA CAG GTC TCA ACA AAC CTA GGA AAA GTG GCG AGC AAA TGT TGT AAA CAT (1430)

441 pro glu ala lys arg met pro cys ala glu asp tyr leu ser val val leu asn gln leu cys val leu his qly lys thr pro val ser 470
 CCT GAA CCA AAA ACA ATG CCC TGT GCA GAA GAC TAT CTA TCC GTC GTC AAC CAG TTA TGT GTC TTC CAT CAG AAA ACG CCA GTA AGT (1520)

471 asp arg val thr lys cys cys thr glu ser leu val asn arg arg pro cys phe ser ala leu glu val asp glu thr thr val pro lys 500
 CAC ACA GTG ACC AAA TCC TCC ACA GAA TCC TTG GTG AAC ACG CCA CCA TCC TTT TCA CCT CTG GAA GTC GAT GAA ACA TAC GTT CCC AAA (1610)

501 glu phe asn ala glu thr phe thr phe his ala asp ile cys thr leu ser glu lys glu arg qln ile lys lys qln thr ala leu val 530
 CAG TTT AAT CCT GAA ACA TTC ACC TTC CAT CCA GAT ATA TCC ACA CTT TCT CAG AAG CAG ACA CAA ATC AAG AAA CAA ACT GCA CTT GTT (1700)

5
10
15
20
25
30
35

531
glu leu val lys his lys pro lys ala thr lys glu gln leu lys ala val met asp phe ala phe val glu lys oys lys
GAG CTC GTG AAA CAC AAG CCC AAG GCA ACA AAA GAG CAA CTG AAA GCT GTT ATG CAT CAT TTC GCT TTT GTA GAG AAG TGC TGC AAG (1790)

540
550
558 559 560

561
ala asp asp lys glu thr oys phe ala glu gln qly lys leu val ala ala ser gln ala ala leu qly leu ter
CCT GAC GAT AAG GAG ACC TGC TTT GCC GAG GAG GGT AAA AAA CTT GTT CCT GCA AGT CAA GCT GCC TTA GCC TTA TAA CATCACATTTAAAG (1883)

567
570
580
ter

581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600

601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700

701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800

801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900

901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000

1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
1021
1022
1023
1024
1025
1026
1027
1028
1029
1030
1031
1032
1033
1034
1035
1036
1037
1038
1039
1040
1041
1042
1043
1044
1045
1046
1047
1048
1049
1050
1051
1052
1053
1054
1055
1056
1057
1058
1059
1060
1061
1062
1063
1064
1065
1066
1067
1068
1069
1070
1071
1072
1073
1074
1075
1076
1077
1078
1079
1080
1081
1082
1083
1084
1085
1086
1087
1088
1089
1090
1091
1092
1093
1094
1095
1096
1097
1098
1099
1100

1101
1102
1103
1104
1105
1106
1107
1108
1109
1110
1111
1112
1113
1114
1115
1116
1117
1118
1119
1120
1121
1122
1123
1124
1125
1126
1127
1128
1129
1130
1131
1132
1133
1134
1135
1136
1137
1138
1139
1140
1141
1142
1143
1144
1145
1146
1147
1148
1149
1150
1151
1152
1153
1154
1155
1156
1157
1158
1159
1160
1161
1162
1163
1164
1165
1166
1167
1168
1169
1170
1171
1172
1173
1174
1175
1176
1177
1178
1179
1180
1181
1182
1183
1184
1185
1186
1187
1188
1189
1190
1191
1192
1193
1194
1195
1196
1197
1198
1199
1200

1201
1202
1203
1204
1205
1206
1207
1208
1209
1210
1211
1212
1213
1214
1215
1216
1217
1218
1219
1220
1221
1222
1223
1224
1225
1226
1227
1228
1229
1230
1231
1232
1233
1234
1235
1236
1237
1238
1239
1240
1241
1242
1243
1244
1245
1246
1247
1248
1249
1250
1251
1252
1253
1254
1255
1256
1257
1258
1259
1260
1261
1262
1263
1264
1265
1266
1267
1268
1269
1270
1271
1272
1273
1274
1275
1276
1277
1278
1279
1280
1281
1282
1283

9. Nucleotide sequence coding for the pre pro human serum albumin, said nucleotide sequence is as follows:

```

5
10
15
20
25
30
35

-18      -10      -1
p r o      p r o      p r o
Met lys trp val tlu phe lle ser leu leu phe leu phe ser
ATG AAC TGG GTA ACC TTT ATT TCC CTT CTT TTT CTC TTT ACC (30)

ser ala tyr ser arg gly val phe arg arg asp ala his lys ser glu val ala his arg phe lys asp leu gly glu asn phe lys
TCG CCT TAT TCC AGG CGT GTG TTT CGT CCA CAT CCA CAC MAG AGT CAG GTT CCT CAT CGG TTT AAA CAT TTG CCA CAA GAA AAT TTC AAA (170)

21      30      34      40      50
ala leu val leu ile ala phe ala gln tyr leu gln gln cys pro phe glu asp his val lys leu val asn glu val thr glu phe ala
CCC TTC GTG TTG ATT GCC TTT GCT CAG TAT CTT CAG CAG TGT CCA TTT GAA CAT CAT GTA AAA TTA GTG AAT GAA GTA ACT CAA TTT CCA (260)

51      60      62      70      75      80
lys thr cys val ala asp glu ser ala glu asn cys asp lys ser leu his thr leu phe gly asp lys leu cys thr val ala thr leu
AAA ACA TGT GTT GCT CAT CAG TCA GCT CAG TAT TGT GAC AAA TCA CTT CAT ACC CTT TTT CCA CAC AAA TTA TCC ACA CTT CCA ACT CTT (350)

81      90      91      100      101      110
arg glu thr tyr gly glu met ala asp cys cys ala lys gln glu pro gly arg asn glu cys phe leu gln his lys asp asp asn pro
CGT GAA ACC TAT GGT CAA ATG GCT CAC TCC TGT GCT GCA AAA CAA CAA CCT CGG ACA AAT GAA TCC TTC TTG CAA CAC AAA CAT GAC AAC CCA (440)

111      120      124      130      140
asn leu pro arg leu val arg pro glu val asp val met cys thr ala phe his asp asn glu thr phe leu lys lys tyr leu try
AAC CTC CCC CCA TTG GTG ACA CCA CAG GTT CAT GTG ATG TGC ACT GCT TTT CAT GAC AAT GAA CAG ACA TTT TTG AAA AAA TAC TTA TAT (530)

141      150      160      168      169      170
glu ile ala arg arg his pro tyr phe tyr ala pro glu leu leu phe phe ala lys arg tyr lys ala ala phe thr glu cys cys gln
CAA ATT GCC ACA ACA CAT CCT TAC TTT TAT GCC CCG GAA CTC CTT TTC TTT GCT AAA AGG TAT AAA GCT GCT TTT ACA GAA TGT TGC CAA (620)

171      177      180      190      200
ala ala asp lys ala ala cys leu leu pro lys leu asp glu leu arg asp glu gly lys ala ser ser ala lys gln arg leu lys cys
GCT GCT GAT AAA GCT GCC TGC CTG TTG CCA MAG CTC GAT GAA GCG MAG GCT TCG TCT GCC AA CAG ACA CTC MAG TGT (710)

201      210      220      230
ala ser leu gln lys phe gly glu arg ala phe lys ala trp ala val ala arg leu ser gln arg phe pro lys ala ala phe ala glu
CCC AGT CTC CAA AAA TTT GCA GAA ACA GCT TTC AAA GCA TGG GCA GTA GCT CCC CTG AGC CAG ACA TTT CCC AAA GCT CAG TTT GCA GAA (800)

```


231 val ser lys leu val thr asp leu thr lys val his thr glu oys oys his gly asp leu leu glu oys ala asp asp arg ala asp leu 260
 GTT TCC AAG TTA GTG ACA CAT CTT ACC AAA GTC CAC ACC GAA TCC TCC CAT GCA CAT CTT GCT GAA TGT GCT CAT CAC AGC GCG CAC CTT (890)

261 ala lys tyr ile oys glu asn gln asp ser ile ser ser lys leu lys glu oys oys qiu lys pro leu leu glu lys ser his oys ile 289 290
 CCC AAG TAT ATC TGT GAA AAT CAA CAT TCC ATC TCC AGT AAA CTG AAG GAA TCC TCT GAA AAA CCT CTG TTG CAA AAA TCT CAC TCC ATT (980)

291 ala glu val glu asn asp glu met pro ala asp leu pro ser leu ala ala asp phe val glu ser lys asp val oys lys asn tyr ala 320
 CCG CAA GTG CAA AAT CAT CAG ATG CCT GCT CAC TTG CCT TCA TTA CCT GCT CAT TTT GCT CAA AGT AAG CAT GCT TGC AAA AAC TAT GCT (1070)

321 glu ala lys asp val phe leu glu met phe leu tyr glu tyr ala arg arg his pro asp tyr ser val val leu leu leu arg leu ala 350
 CAG CCA AAG CAT GTC TTC TTG GCG ATG TTT TTG TAT GAA TAT GCA AGA AGC CAT CCT CAT TAC TCT GTC GTG CTG CTG CCA CTT GCC (1160)

351 lys thr tyr glu thr leu glu lys oys ala ala asp pro his qiu oys tyr ala lys val phe asp qiu phe lys pro leu 380
 AAG ACA TAT CAA ACC ACT CTA GAG AAG TCC TGT GCG CCT GCA CAT CCA TCC TAT GCG AAA GTG TTC CAT GAA TTT AAA CCT CCT (1250)

381 val glu glu pro gln asn leu ile lys gln asn oys glu leu phe qiu gln leu qiu glu tyr lys phe gln asn ala leu leu val arg 410
 GTG CAA GAG CCT CAG AAT TTA ATC AAA CAA AAT TGT GAG CTT TTT GAG CAG CTT CCA CAG TAC AAA TTC CAG AAT CCG CTG TTA CTT CCG (1340)

411 tyr thr lys lys val pro gln val ser thr pro thr leu val glu val ser arg asn leu qiu lys val qiu lys val qiu lys ser lys oys lys his 440
 TAC ACC AAG AAA GTA CCC CAA GTG TCA ACT CCA ACT CTT GTA GAG GTC TCA ACA AAC CTA CCA AAA GTG CCG ACC AAA TGT TGT AAA CAT (1430)

441 pro glu ala lys arg met pro oys ala glu asp tyr leu ser val val leu asn gln leu oys val leu his qiu lys thr pro val ser 470
 CCT CAA GCA AAA ACA ATG CCC TGT GCA GAA CAC TAT CTA TCC GTG GTC CTG AAC CAG TTA TGT GTG TTG CAT CAG AAA ACG CCA GTA ACT (1520)

471 asp arg val thr lys oys oys thr glu ser leu val asn arg arg pro oys phe ser ala leu glu val asp qiu thr tyr val pro lys 500
 CAC ACA GTC ACC AAA TCC TCC ACA GAA TCC TTG GTG AAC AGG CGA CCA TCC TTT TCA CCT CTG CAA GTC CAT GAA ACA TAC GTT CCC AAA (1610)

501 glu phe asn ala glu thr phe thr phe his ala asp ile oys thr leu ser glu lys glu arg qiu ile lys lys qiu thr ala leu val 530
 GAG TTT AAT GCT GAA ACA TTC ACC TTC CAT CCA GAT ATA TCC ACA CTT TCT GAG AAG GAG ACA CAA ATC AAG AAA CAA ACT CCA CTT GTT (1700)

5
10
15
20
25
30
35

531 glu leu val lys his lys pro lys ala thr lys glu gln leu lys ala val met asp phe ala ala phe val glu lys cys lys
CAG CTC GTG AAA CAC AAG CCC AAG GCA ACA AAA GAG CAA CAG CAA CTG AAA GCT GTT ATG GAT GAT TTC GCT GCT TTT GTA GAG AAG TGC TGC AAG (1790)

540
550 558 559 560
561 ala asp asp lys glu thr cys phe ala glu glu gln lys lys leu val ala ala ser gln ala ala leu gly leu ter
CCT CAC GAT AAG CAG ACC TGC TTT GCC CAG CAG GGT AAA AAA CTT GTT GCT GCA AGT CAA GCT GCC TTA GGC TTA TAA CATCACATTAAAG (1883)

567 570 580
ter ter
CATCTACGCTACCATGAGATAAGAGAAATGAGCATCAAGAGCTTATTCACTGTTTTCTTTTTTCGTTGCTGTAAGCCACACCCCTGCTCTAAAAACATAAATTTCTTTAA (2002)

TCATTTGCCCTCTTTTCTGCTGCTTCAATTAAATAAAATGCAAGCATCTAA..... 20AA (2078)

10. A nucleotide sequence according to any of claims 6 to 9, in essentially pure form.
11. A DNA transfer vector comprising a nucleotide sequence as defined in claim 5.
- 5 12. A DNA transfer vector according to claim 11, transferred to and replicated in a micro-organism.
13. A DNA transfer vector according to claim 12, which is a plasmid.
14. A DNA transfer vector according to claim 13,
10 wherein the plasmid is pBR322 or YEp6.
15. A process for preparing human serum albumin, which comprises culturing a micro-organism according to claim 5.
16. A DNA transfer vector according to any of
15 claims 12 to 14, or a process according to claim 15, wherein the micro-organism is a bacterium or yeast.
17. A vector or process according to claim 16, wherein the bacterium or yeast is E. coli or Saccharomyces cerevisiae.

Restriction Endonuclease Map of Human Serum Albumin cDNA Clones

PHA36

